November 18, 2024

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York

> Re: *The New York Times Company v. Microsoft Corp.* No.: 23-cv-11195; *Daily News, LP v. Microsoft Corp.*: No. 1:24-cv-3285; *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*: No. 1:24-cv-04872: Dispute Regarding Additional OpenAI ESI Custodians
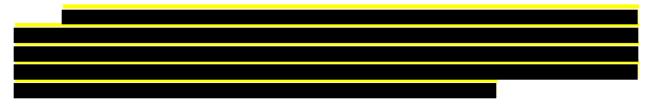
Dear Magistrate Judge Wang:

Plaintiffs The New York Times Company ("The Times"), Daily News, LP et al. ("Daily News"), and the Center for Investigative Reporting ("CIR") (collectively, the "News Plaintiffs") request an order compelling OpenAI to add four additional custodians: **Michael Lampe**, **Suchir Balaji**, **Peter Deng**, and **Ethan Stock**, for the reasons set forth below. The News Plaintiffs also reassert their request for the eight custodians in their August 12 motion to compel, including because hit counts provided since the last hearing support the News Plaintiffs' request for these custodians. *See* Dkt. 204 (motion); Dkt. 211 (opposition).[1]

I.   The Four New Custodians

The News Plaintiffs have continued reviewing OpenAI's document productions and, based on that review and other information, request four additional ESI custodians from OpenAI.[2]
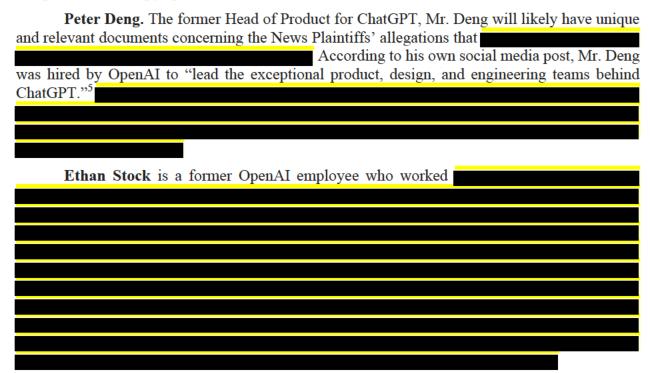
Michael Lampe. A member of OpenAI's technical staff since before the release of ChatGPT in November 2022, Mr. Lampe has worked directly on issues concerning how Defendants' models and products rely upon copyrighted content. For this reason, the court overseeing the class action in the Northern District of California recently ordered that Mr. Lampe be added as a custodian in that case. *See Tremblay v. OpenAI*, 23-cv-03223, Dkt. 191. The California order describes how Mr. Lampe "reportedly oversaw relevant projects related to testing the language models' understanding of whether a work is copyright-protected or in the public domain; and, that he appears to have developed the source code OpenAI has used to prevent the models from regurgitating copyright-protected material." *Id.*

[REDACTED]

---

[1] Those eight custodians are: James Betker, William Fedus, Daniel, Kokotajlo, Brad Lightcap, Ryan Lowe, William Saunders, Ilya Sutskever, and Wojciech Zaremba.

[2] The News Plaintiffs sent OpenAI an email on November 13 requesting these four custodians and providing their bases for seeking them, but OpenAI has not yet responded, despite repeated follow-up. *See* Ex 1 (email correspondence).

Suchir Balaji. Mr. Balaji is a former OpenAI employee who worked on post-training for ChatGPT and pretraining for the GPT-4 model.[3] He will have unique and relevant documents supporting The News Plaintiffs' allegations of willful infringement, among other issues. New York Times FAC ¶¶ 124-26. Mr. Balaji recently (and publicly) expressed his view that OpenAI committed copyright infringement through ChatGPT. A personal blog post describes how "the process of training a generative model involves making copies of copyrighted data" and why "training on copyrighted data without a [] licensing agreement is also a type of market harm, because it deprives the copyright holder of a source of revenue."[4] His post emphasizes that it's "pretty obvious" that ChatGPT generates "market harms" because it "produc[es] substitutes" that compete with the copyrighted content on which ChatGPT was trained.

Peter Deng. The former Head of Product for ChatGPT, Mr. Deng will likely have unique and relevant documents concerning the News Plaintiffs' allegations that ███████████████████ ███████████████████████████ According to his own social media post, Mr. Deng was hired by OpenAI to "lead the exceptional product, design, and engineering teams behind ChatGPT."[5]███

Ethan Stock is a former OpenAI employee who worked ███

II.   The Eight Previously Disputed Custodians

The News Plaintiffs also stand by their request for OpenAI to add the eight disputed custodians from The Times's August 12 motion to compel. Dkt. 204 (motion); Dkt. 211 (opposition). In total, the News Plaintiffs seek twelve additional OpenAI custodians, for a total of 38 custodians (on top of the already agreed to 26 custodians). *See Fort Worth Employees' Ret. Fund v. J.P. Morgan Chase & Co.*, 297 F.R.D. 99, 106 (S.D.N.Y. 2013) (ordering 18 additional custodians, on top of 42 previous custodians, for a total of 60 custodians).

Following the last hearing, OpenAI produced a de-duplicated hit count for these custodians which showed 122,926 potentially relevant documents, an amount more than proportional to the needs of the case. *See Fort Worth*, 297 F.R.D. at 103 (requiring the defendant to apply additional

---

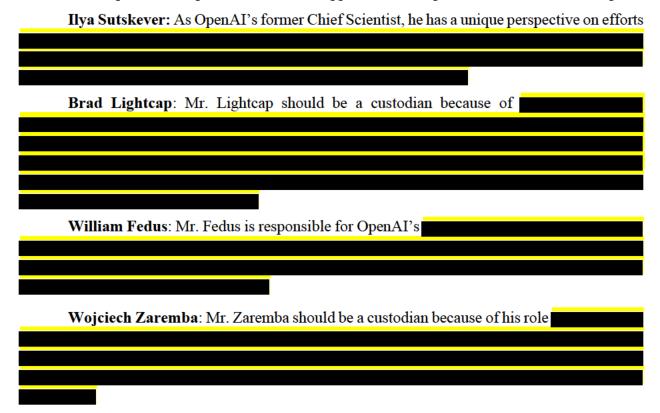[3] https://www.linkedin.com/in/suchirbalaji/.

[4] https://suchir.net/fair_use.html.

[5] https://www.jagranjosh.com/general-knowledge/who-is-peter-deng-1685602455-1.

search terms, notwithstanding defendant's argument that it had already reviewed 875,000 documents).

Moreover, this hit count will decrease because it was based on the News Plaintiffs' *prior* search term proposal, which yielded approximately 630,000 unique hits for the 26 agreed upon custodians. Consistent with the Court's ruling, the News Plaintiffs made a narrower proposal to yield around 500,000 documents for the current set of 26 agreed custodians. Dkt. 304 (ruling that "the 'total hit count' will be capped at or around 500,000 for the current set of agreed custodians, without regard to the number of Disputed Custodians that later may be added"). The News Plaintiffs' narrowed proposal will shrink the hit counts for the eight disputed custodians as well, likely to around 100,000 unique hits if the proportion reduced is similar to that of the original custodians. Assuming OpenAI's reported 25% responsiveness rate continues, Dkt. 304, this means only an additional 25,000 documents will be produced. That number is more than proportional, particularly given the scope of the products and conduct at issue.

Also, since the prior conference on October 30, OpenAI has produced additional documents from custodians' files, which the News Plaintiffs are diligently reviewing. Documents in these recent productions provide additional support for the disputed custodians. For example:

Ilya Sutskever: As OpenAI's former Chief Scientist, he has a unique perspective on efforts ██ ██ ████████████████████████████████████

Brad Lightcap: Mr. Lightcap should be a custodian because of ████████ ██ ██ ██ ██ ████████████████

William Fedus: Mr. Fedus is responsible for OpenAI's ████████████ ██ ██ ████████████████

Wojciech Zaremba: Mr. Zaremba should be a custodian because of his role ████ ██ ██ ██ ████████

The News Plaintiffs respectfully request a ruling on all twelve of the disputed custodians so the parties can conclude custodian and search terms negotiations and promptly move forward to depositions.

Respectfully,

/s/ Ian B. Crosby

Ian B. Crosby
Susman Godfrey L.L.P.

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

/s/ Matt Topic
Matt Topic
Loevy & Loevy

cc: All Counsel of Record (via ECF)